**Hadoop实践作业**

用hadoop的streaming mapreduce或者pig，或者开发下面几个程序，完成下面的要求：

1. 统计首字母分别为a~z的单词数目，以及首字母为a~z的单词平均长度；
2. 统计尾字母分别为a~z的单词数目，以及尾字母为a~z的单词的平均长度；
3. 按照单词首字母和尾字母的组合统计单词数目，以及各个组合里面单词平均长度；（举个例子，单词为apple,那么首字母为a，尾字母为e, apple这个单词就要归属到首字母为a尾字母为e的那一组。理论上组合有26\*26种，但有的组合可能并没有单词）

统计的数据：hadoop\_input, 见附件

作业上交的要求：

1. 以组为单位 (我已经跟陈老师沟通过了,咱们就按照自己的学习小组交即可)；
2. 需要上交代码，运行方式，以及最终结果；
3. 作业上交时写清楚组里的人员名单；
4. 提交日期:1月15日16点, 小组形式上交,发给班长/副班长! 最后由班长汇总,1月16日16点前发到我qq邮箱,大家不要逾期提交,最后学期末作业都需要提交存档的!